

Proteome-wide dataset supporting functional study of tyrosine kinases in breast cancer

Article (Published Version)

Angelopoulos, Nicos, Stebbing, Justin, Xu, Yichen, Giamas, Georgios and Zhang, Hua (2016) Proteome-wide dataset supporting functional study of tyrosine kinases in breast cancer. Data in Brief, 7. pp. 740-746. ISSN 2352-3409

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/61059/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data article

Proteome-wide dataset supporting functional study of tyrosine kinases in breast cancer

Nicos Angelopoulos^a, Justin Stebbing^a, Yichen Xu^a,
Georgios Giamas^{a,b,**}, Hua Zhang^{a,*}^a Department of Surgery and Cancer, Imperial College London, Hammersmith Hospital Campus, Du Cane Road, London W12 0NN, UK^b School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9RQ, UK

ARTICLE INFO

Article history:

Received 10 July 2015

Received in revised form

24 February 2016

Accepted 4 March 2016

Available online 14 March 2016

Keywords:

Breast cancer

Cell signaling

Proteomics

SILAC

Tyrosine kinases

ABSTRACT

Tyrosine kinases (TKs) play an essential role in regulating various cellular activities and dysregulation of TK signaling contributes to oncogenesis. However, less than half of the TKs have been thoroughly studied. Through a combined use of RNAi and stable isotope labeling with amino acids in cell culture (SILAC)-based quantitative proteomics, a global functional proteomic landscape of TKs in breast cancer was recently revealed highlighting a comprehensive and highly integrated signaling network regulated by TKs (Stebbing et al., 2015) [1]. We collate the enormous amount of the proteomic data in an open access platform, providing a valuable resource for studying the function of TKs in cancer and benefiting the science community. Here we present a detailed description related to this study (Stebbing et al., 2015) [1] and the raw data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the identifier PXD002065.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

* Corresponding author.

** Corresponding author at: Department of Surgery and Cancer, Imperial College London, Hammersmith Hospital Campus, Du Cane Road, London W12 0NN, UK.

E-mail address: h.zhang10@imperial.ac.uk (H. Zhang).<http://dx.doi.org/10.1016/j.dib.2016.03.024>2352-3409/© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Specifications Table

Subject area	Biology
More specific sub- ject area	Cancer biology
Type of data	Mass spectrometry (MS) data and annotations
How data was acquired	LTQ-Orbitrap Velos MS (Thermo Scientific)
Data format	Raw and MaxQuant output data
Experimental factors	RNAi of the 65 expressed TKs in breast cancer cells (MCF-7)
Experimental features	Cells cultured in media with different SILAC labeling: 'medium' L-[¹³ C ₆] arginine (R6) and L-[² H ₄] lysine (K4), and 'heavy' L-[¹³ C ₆ , ¹⁵ N ₄] arginine (R10) and L-[¹³ C ₆ , ¹⁵ N ₂] lysine (K8), and 'light' L-[¹² C ₆ , ¹⁴ N ₄] arginine (R0) and L-[¹² C ₆ , ¹⁴ N ₂] lysine (K0)
Data source location	London, UK
Data accessibility	The raw data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the identifier PXD002065. http://www.ebi.ac.uk/pride/archive/projects/PXD002065

Value of the data

- This is the first dataset to examine the entirety of a TK-modulated proteome in breast cancer.
- For all the 65 TKs, from 27,000 non-redundant peptide sequences, a SILAC-based proteome encompassing more than 2000 distinguishable and unambiguously identified proteins was identified and quantified with a minimum of two peptides with a false discovery rate (FDR) of 1%.
- Functional analysis shows that TK-regulated proteome are involved in a broad variety of cellular activities including immune system process, metabolic process, growth, cell cycle, transcription, and apoptosis, whose deregulation can contribute to carcinogenesis.
- Of note, instead of structural homology of TKs, our dataset highlights similarity in their biologic function and shows a detailed portrait of their signaling networks.
- Our TK-proteome dataset provides an enormous resource for functional study of TKs in cancer, particularly for studying the downstream effects and associated signaling modulated by TKs.

1. Data

This dataset collates the comprehensive proteomic data generated through a SILAC-based quantitative MS approach upon silencing the 65 tyrosine kinases expressed in MCF7 breast cancer cells [1]. In addition to the raw data files deposited in the Pride database [2,3], here we also provide a new quantification master-file that provides easy access to the TK SILAC dataset (Supplementary Table 1). Each row provides information about the relative quantification of a protein in all knockdowns. The data are organized by protein, with the HGNC symbol and Entrez identifier of corresponding genes for cross references. Users of this dataset can easily access and cross reference our dataset to external databases and to their in-house experimental data.

2. Experimental design, materials and methods

2.1. Experimental design

To acquire the TK-regulated proteome and associated signaling dynamics, a combined approach of RNAi with SILAC-based quantitative proteomics was employed in breast cancer cells MCF-7. Subsequently, bioinformatics analyses were implemented to reclassify the TKs family and to characterize the associated functional portrait (Fig. 1).

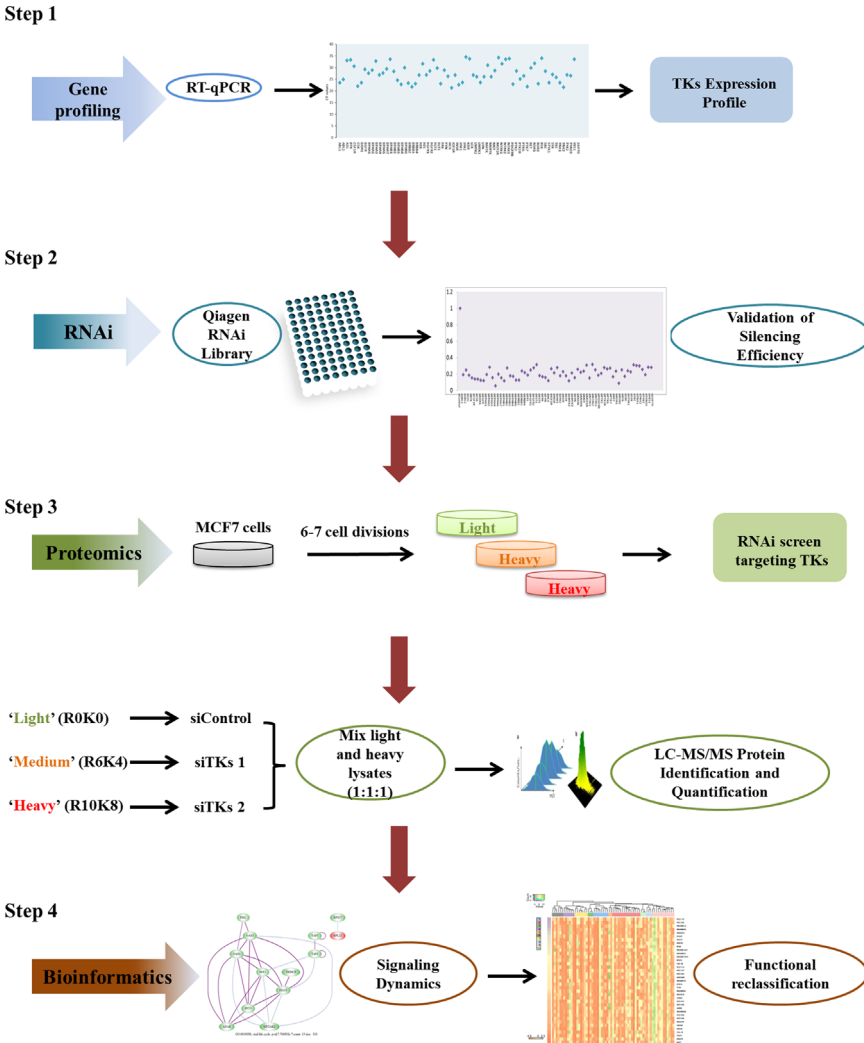


Fig. 1. Workflow of the experimental design. Firstly, the gene expression of the 90 TKs was profiled by RT-qPCR in MCF7 cells. Next, the knock-down efficiency of a certified RNAi library comprising two individual siRNAs against TKs was confirmed by RT-PCR and western blotting. Subsequently, MCF7 cells grown in either R0K0 'light', or R6K4 'medium', or R10K8 'heavy' were treated with siControl or verified siRNAs targeting the TKs for 72 h. Protein samples were then harvested and analyzed by SILAC-based MS quantitative proteomics. Further bioinformatic analyses were conducted to characterize the signaling dynamics, to establish the associated functional portrait and to reclassify the family of TKs.

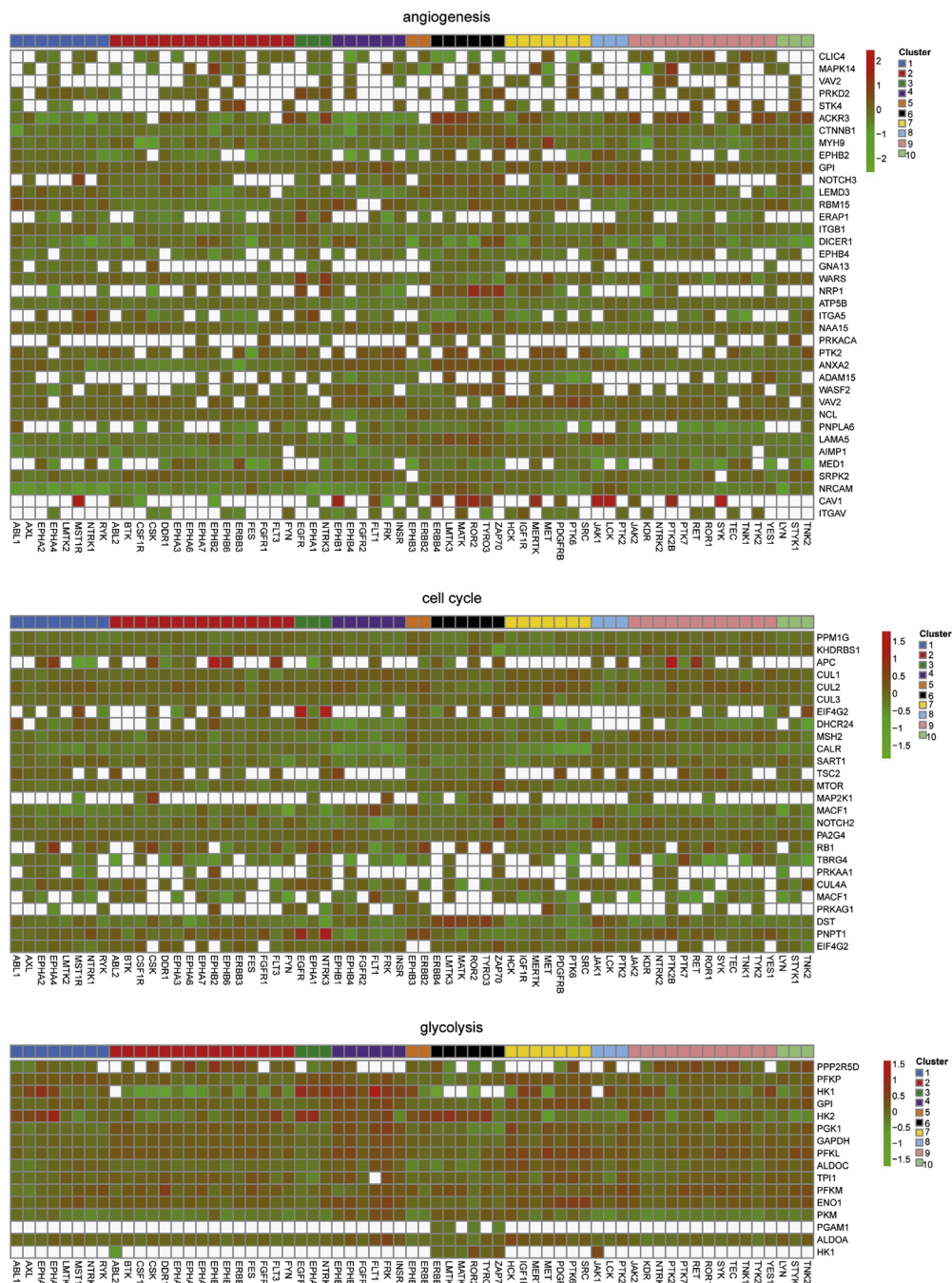


Fig. 2. Heatmap of the altered proteomic quantifications involved in studied cancer hallmarks (angiogenesis, cell cycle, glycolysis). Presented here are log₂ values of normalized fold changes against control for the significantly up- (red) or down-regulated (green) proteins (Significant B test $P < 0.05$) upon silencing TKs in our dataset. Ten distinctive clusters were shown color-coded as in our original publication.

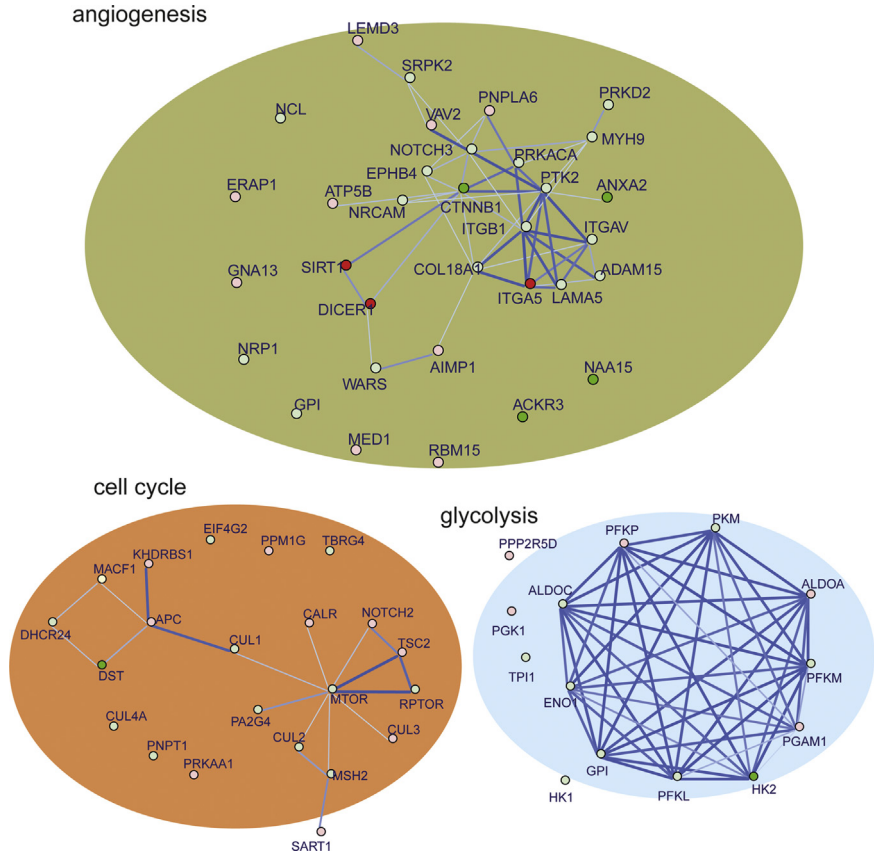


Fig. 3. Representatives of defined functional networks in angiogenesis, cell cycle, and glycolysis. The functional networks were generated using GO analysis combined with the STRING platform. Proteins in red are up-regulated, whereas green indicates down-regulation.

Here, to show how our data could be more valuable to the scientific community, we sought to further investigate the biological significance of our TKs proteomic dataset in studying the hallmarks of cancer [4]. GO terms of cancer hallmarks, such as angiogenesis, cell cycle, glycolysis, motility and apoptosis, were searched and a list of corresponding proteins/genes was identified. We then first examined the overall effect TKs on each GO family by plotting a heatmap of all the term proteins that were quantified in our dataset. Representative heatmaps showing quantified proteins involved in angiogenesis, cell cycle, glycolysis are presented based on our classified clusters (Fig. 2 and Supplementary Fig. 1). These findings highlight the involvement of TKs in the hallmarks of cancer. Of note, our analysis was able to identify the top targets that were significantly modulated by TKs in each category, including the Lemur Tyrosine Kinase 3 (LMTK3) in cell migration [5]. In addition, to establish the functional protein–protein interaction networks in these hallmarks of cancer, we integrated the GO analyses with the STRING interaction database. We illustrated the STRING network for the differential proteins within each represented hallmark of cancer, including in angiogenesis, cell cycle, and glycolysis (Fig. 3). These findings reinforce the essential role of TKs in regulating hallmarks of cancer and highlight the complexities of TK-modulated cellular signaling.

2.2. SILAC labeling and RNAi screening

DMEM medium deficient in arginine (R) and lysine (K) was supplemented with stable isotope-encoded arginine and lysine as previously described [6,7]. Briefly, medium were provided with either 'light' L-[$^{12}\text{C}_6$, $^{14}\text{N}_4$] arginine (R0) and L-[$^{12}\text{C}_6$, $^{14}\text{N}_2$] lysine (K0), or 'medium' L-[$^{13}\text{C}_6$] arginine (R6) and L-[$^2\text{H}_4$] lysine (K4), or 'heavy' L-[$^{13}\text{C}_6$, $^{15}\text{N}_4$] arginine (R10) and L-[$^{13}\text{C}_6$, $^{15}\text{N}_2$] lysine (K8) from Dundee Cell Products Ltd. (Dundee, UK). MCF7 cells were grown in these DMEM mediums for at least seven divisions prior to experiments. The human Tyrosine Kinase family siRNAs library (Qiagen FlexiPlate) comprising two individual verified siRNAs was utilized (final concentration 40 nM) for 72 h ([Supplementary Table 2](#)).

2.3. Protein digestion and peptide fractionation

Prior to digestion, equal amounts of lysates from unlabeled and labeled samples were mixed. Next, samples were reduced in 10 mM DTT and alkylated in 50 mM Iodoacetamide followed by boiling in loading buffer, and then separated by one-dimensional SDS-PAGE (4–12% Bis-Tris Novex mini-gel, Invitrogen) and visualized by colloidal Coomassie staining (Novex, Invitrogen). The entire protein gel lanes were excised and cut into 10 slices each, with each gel slice subjected to in-gel digestion with trypsin overnight at 37 °C. The resulting tryptic peptides were extracted by formic acid (1%) and acetonitrile (CH_3CN), lyophilized in a speedvac and resuspended in 1% formic acid.

2.4. Mass spectrometry

Trypsin-digested peptides were separated using an Ultimate 3000 RSLC (Thermo Scientific) nanoflow LC system. On average 0.5 μg was loaded with a constant flow of 5 $\mu\text{l}/\text{min}$ onto an Acclaim PepMap100 nanoViper C18 trap column (100 μm inner-diameter, 2 cm; Thermo Scientific). After trap enrichment, peptides were eluted onto an Acclaim PepMap RSLC nanoViper, C18 column (75 μm , 15 cm; Thermo Scientific) with a linear gradient of 2–40% solvent B (80% acetonitrile with 0.08% formic acid) over 65 min with a constant flow of 300 nl/min . The HPLC system was coupled to a linear ion trap Orbitrap hybrid mass spectrometer (LTQ-Orbitrap Velos, Thermo Scientific) via a nano electrospray ion source (Thermo Scientific). The spray voltage was set to 1.2 kV, and the temperature of the heated capillary was set to 250 °C. Full-scan MS survey spectra (m/z 335–1800) in profile mode were acquired in the Orbitrap with a resolution of 60,000 after accumulation of 1,000,000 ions. The fifteen most intense peptide ions from the preview scan in the Orbitrap were fragmented by collision-induced dissociation (normalized collision energy, 35%; activation Q, 0.250; and activation time, 10 ms) in the LTQ Orbitrap after the accumulation of 10,000 ions. Maximal filling times were 1000 ms for the full scans and 150 ms for the MS/MS scans.

2.5. Proteome quantification

The raw mass spectrometric data files were collated using MaxQuant [8] and the Andromeda search engine software [9]. Peptide ratios were calculated for each arginine- and/or lysine-containing peptide as the peak area of labeled arginine/lysine divided by the peak area of non-labeled arginine/lysine for each single-scan mass spectrum. Data are normalized using 1/median ratio value for each identified protein group per labeled sample. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [2] via the PRIDE partner repository with the dataset identifier PXD002065. The corresponding labeling of the respective silencing of the TKs in the dataset is shown in [Supplementary Table 3](#). Annotated spectra for all results can be accessed using MS-Viewer [10] via the links provided from [Supplementary Table 4](#).

2.6. Bioinformatics analysis

The bioinformatics analyses were developed in R [11] and SWI-Prolog [12], using Real [13] for linking the two systems. The hierarchical clustering of proteomics data was performed based on

calculated distances by R's `hclust` function. To visualize the TK-modulated proteomics, the heatmap of quantified values exhibiting the overall pattern of regulation was displayed. Significance B test was used to characterize the most significantly regulated proteins after silencing of TKs ($p < 0.05$) [8]. For analyzing Gene ontology (GO) [14] and STRING [15] network, GO, protein–protein interaction searches and identifier mapping was via the `bio_db` library with primary data provided by GO, HGNC and STRING. SWI-Prolog was used for constructing the graphs and R library `igraph` was deployed for displaying the graphs. Protein–protein interactions were provided by String version 10. The HGNC and GO databases were accessed on 2015/9/28.

Acknowledgments

This work is supported by Action Against Cancer, UK.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2016.03.024>.

References

- [1] J. Stebbing, H. Zhang, Y. Xu, et al., Characterization of the tyrosine kinase-regulated proteome in breast cancer by combined use of RNA interference (RNAi) and stable isotope labeling with amino acids in cell culture (SILAC) quantitative proteomics, *Mol. Cell. Proteom.* 14 (9) (2015) 2479–2492.
- [2] J.A. Vizcaino, E.W. Deutsch, R. Wang, et al., ProteomeXchange provides globally coordinated proteomics data submission and dissemination, *Nat. Biotechnol.* 32 (3) (2014) 223–226.
- [3] J.A. Vizcaino, R.G. Cote, A. Csordas, et al., The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013, *Nucleic Acids Res.* 41 (2013) D1063–D1069 (Database issue).
- [4] D. Hanahan, R.A. Weinberg, Hallmarks of cancer: the next generation, *Cell* 144 (5) (2011) 646–674.
- [5] Y. Xu, H. Zhang, L.C. Lit, et al., The kinase LMTK3 promotes invasion in breast cancer through GRB2-mediated induction of integrin beta(1), *Sci. Signal.* 7 (330) (2014) ra58.
- [6] H. Zhang, N. Angelopoulos, Y. Xu, et al., Proteomic profile of KSR1-regulated signalling in response to genotoxic agents in breast cancer, *Breast Cancer Res. Treat.* 151 (3) (2015) 555–568.
- [7] H. Zhang, Y. Xu, A. Filipovic, et al., SILAC-based phosphoproteomics reveals an inhibitory role of KSR1 in p53 transcriptional activity via modulation of DBC1, *Br. J. Cancer* 109 (10) (2013) 2675–2684.
- [8] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat. Biotechnol.* 26 (12) (2008) 1367–1372.
- [9] J. Cox, N. Neuhauser, A. Michalski, R.A. Scheltema, J.V. Olsen, M. Mann, Andromeda: a peptide search engine integrated into the MaxQuant environment, *J. Proteome Res.* 10 (4) (2011) 1794–1805.
- [10] P.R. Baker, R.J. Chalkley, MS-viewer: a web-based spectral viewer for proteomics results, *Mol. Cell. Proteom.* 13 (5) (2014) 1392–1396.
- [11] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, 2014.
- [12] J. Wielemaker, T. Schrijvers, M. Triska, T. Lager, SWI-Prolog, *Theory Pract. Log. Program.* 12 (1–2) (2012) 67–96.
- [13] N. Angelopoulos, V.S. Costa, J. Azevedo, J. Wielemaker, R. Camacho, L. Wessels, Integrative functional statistics in logic programming, in: *Proceedings of Practical Aspects of Declarative Languages*, Rome, Italy, 2013, pp. 190–205.
- [14] M. Ashburner, C.A. Ball, J.A. Blake, et al., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, *Nat. Genet.* 25 (1) (2000) 25–29.
- [15] D. Szklarczyk, A. Franceschini, S. Wyder, et al., STRING v10: protein–protein interaction networks, integrated over the tree of life, *Nucleic Acids Res.* 43 (2015) D447–D452 (Database issue).